AD-A264 800

.JTATION PAGE

②

1a. REPORT SECURITY CL	RESTRICTIVE MARKINGS
Unclassified	

2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT
	Approved for public release; distribution
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
	#68

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
Institute for Brain and Neural Systems		Personnel and Training Research Programs Office of Naval Research (Code 1142PT)

6c. ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State and ZIP Code)
Brown University Providence, Rhode Island 02912	800 North Quincy Street Arlington, VA 22217-5000

8a. NAME OF FUNDING / SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
		N00014-91-J-1316

8c. ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO

DTIC ELECTE MAY 25 1993

11 TITLE (Include Security Classification)

Principal Components of Natural Images - an Analytical Solution.

12 PERSONAL AUTHOR(S)
Yong Liu and Harel Shouval

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical Report	FROM _____ TO _____	May 17, 1993	Eight

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Principal Components, Hebbian Synapses, Cortical Receptive
05	08		Fields.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The structure of receptive fields in the visual cortex is believed to be shaped by unsupervised learning. A simple variant of unsupervised learning is the extraction of principal components. In this paper, we derived analytically the form of the principal components of natural images. This derivation relies on results about the covariance matrix of natural images (Field, 1987). Our results predict both the shapes and the phases of the receptive fields. We also compared our results to numerical simulation results (Hancock et al., 1992). Finally the biological relevance of our results is discussed.

93-11611

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	Unclassified

22a NAME OF RESPONSIBLE INDIVIDUAL	22b TELEPHONE (Include Area Code)	22c OFFICE SYMBOL
Dr. Joel Davis	(703) 696-4744	

Principal Components of Natural Images - an Analytic Solution

Yong Liu   **Harel Shouval**

Department of Physics and

Institute for Brain and Neural Systems

Box 1843, Brown University

Providence, R. I., 02912

yong@cns.brown.edu

hzs@cns.brown.edu

December 17, 1992

Abstract

The structure of receptive fields in the visual cortex is believed to be shaped by unsupervised learning . A simple variant of unsupervised learning is the extraction of principal components. In this paper, we derived analytically the form of the principal components of natural images. This derivation relies on results about the covariance matrix of natural images (Field, 1987). Our resluts predict both the shapes and the phases of the receptive fields. We also compared our resluts to numerical simulation results (Hancock et al., 1992). Finally the biological relevance of our results is discussed.

1 Introduction

It is generally believed that the shape of the receptive fields in the visual cortex is determined by some form of unsupervised learning. Hebb's postulate (Hebb, 1949) is the cornerstone of most unsupervised learning models in neural networks. The naive Hebbian rule, however, is unstable. This problem is often overcome by adding constraints to this rule. One such constraint is adding a decay term to this rule. It has been proven that a neuron with Hebbian learning rule plus a proper decay term can perform a principal component extraction (Oja, 1982). Furthermore, a neural network with proper lateral inhibition can perform the extraction of several principal components simultaneously (Sanger, 1989). The computational importance of principal components is that they are the optimal linear projections for minimizing the mean squared reconstruction error (Fukunaga, 1990).

Since the principal components of a set of inputs depend only on their covariance matrix, it is reasonable that given this matrix, they can be calculated analytically. A known example is Linsker's multilayered network (Linsker, 1986; Linsker, 1988), which is trained with random noise. The covariance matrix results from the existence of the Arbor functions between the random input layer and the first layer. This induced correlation was used by several investigators to give an analytic explanation of Linsker's result (Kammen and Yuille, 1988; Yuille et al., 1989; Tang, 1989; MacKay and Miller, 1990a; MacKay and Miller, 1990b) [1].

We believe that modeling the environment with natural scenes is more reasonable than modeling it with random noise or bars or edges, etc., because the environment to which the visual cortex is exposed, through the visual pathway, is composed of natural scenes. When the input environment is not random, the properties of the covariance matrix depend on the nature of the environment, and on the preprocessing performed by the visual pathway. However, in this paper, we have simplified the problem by neglecting the effects of this preprocessing, as a first step towards a more complete mathematical understanding of the receptive field structure of neurons in the visual cortex. Another aspect of this simplification is that we can use our results to explain the simulation results (Hancock et al., 1992), in which the authors used a network of the type proposed by Sanger (Sanger, 1989) to extract principal components from natural scenes. In their simulations, no preprocessing was applied to the images, apart from multiplying the inputs by a smooth Gaussian function, in order to eliminate edge effects.

The nature of the covariance matrix of natural images was investigated by Field (Field, 1987). He found that the spectrum of covariance matrix is proportional to the inverse of the square of the frequency. In section 2, we will derive analytically the form of the principal components of natural images under this assumption.. Rather than taking a smooth Gaussian window as Hancock et. al. did, we assume a circular hard boundary to the receptive fields. We find that the solutions are the Fourier-Bessel functions (Jackson, 1975, etc.). We will show in section 3, that under the assumption that the covariance matrix spectrum has a small non-rotationally symmetric correction, the solutions have a definite phase.

When simulation results (Hancock et al., 1992) are compared to our results, we find a good agreement for the higher eigenvalue solutions, with some deviations for lower eigenvalue solutions. In the discussion we explain how these deviations may come about.

It is clear that the receptive fields obtained here are not identical to the receptive field structures obtained experimentally in the visual cortex. Does this imply that neurons in the first layers of the visual cortex are not principal component analyzers? We will address this issue in the discussion.

---

[1] Their solutions are to an equation with some additional terms to the principal component equation, but it is similar to a principal component equation.

2 The Rotationally Symmetric Solution

The principal components are the eigen-functions of the covariance matrix. Therefore the equation we try to solve is the eigenvalue problem, i.e., the eigen-equation, which has the form

$$\sum_j C_{ij} w_j = \lambda w_i \tag{1}$$

where w_i are eigen-vectors, λ is the eigenvalue, and C_{ij} is the covariance matrix which is defined as $C_{ij} = E[(I_i - E[I_i])(I_j - E[I_j])]$ for input pattern $\{I_i\}$. Since we are dealing with two dimensional space, the index i really denotes a point in the two dimensional space, so it is more convenient to rewrite the covariance matrix in the form $C(\mathbf{r}_i, \mathbf{r}'_j)$. Due to translational invariance, $C(\mathbf{r}_i, \mathbf{r}'_j) = C(\mathbf{r}_i - \mathbf{r}'_j)$. In the continuous limit, the summation will become an integral over $\mathbf{r}'$, thus the eigen-equation becomes

$$\int C(\mathbf{r} - \mathbf{r}') w(\mathbf{r}') d^2 \mathbf{r}' = \lambda w(\mathbf{r}). \tag{2}$$

in which $w(\mathbf{r})$ is the continuous limit of the eigen-vectors w_i.

The Fourier transform (spectrum) of the covariance matrix has the form, $C(\mathbf{k}) = c/\mathbf{k}^2$ where c is a constant (Field, 1987). Hereafter we will set $c = 1$ for convenience. Thus $C(\mathbf{r})$ satisfies $\nabla^2 C(\mathbf{r}) = -\delta(\mathbf{r})$ which can be readily proven by taking Fourier transformation on both side of this equation. Therefore by operating both sides of the eigen-equation 2 with ∇^2, we get

$$\nabla^2 w(\mathbf{r}) = \frac{1}{\lambda} w(\mathbf{r}). \tag{3}$$

In order to find a closed form solution to this equation, we assume that $w(r)$ is non-zero only within a circle of radius a. The justification for this is that the receptive field of a neuron is always of finite size.[2] The equation can be expressed in the polar coordinate, and the solutions are (Jackson, 1975, etc.)

$$w_{mi}(\mathbf{r}) = \begin{cases} J_m(\frac{r}{\sqrt{\lambda_{mi}}}) \begin{cases} cos(m\theta) \\ sin(m\theta) \end{cases} & \text{for } r \le a \\ 0 & \text{for } r > a \end{cases} \tag{4}$$

in which $m = 0, 1, 2, ...$, $J_m(x)$ is the standard Bessel functions, λ_{mi} is the ith root of equation $J_m(a/\sqrt{\lambda}) = 0$, r and θ are the polar coordinate of $\mathbf{r}$. Because the complete solutions of the original eigenvalue problem must be a complete and orthonormal set of functions, the complete set of solutions to the eigenvalue problem is [3]

---

[2]A Gaussian mask used in the numerical simulation (Hancock et al., 1992) is a relaxed version of the boundary condition we have here. Solving the exact equation with a Gaussian mask becomes mathematically more complicated. We should expect that for a Gaussian mask the solutions near the center should be almost identical to the hard boundary solutions, but they would approach zero gradually near the boundary.

[3]Of course the trivial solution that $w^1_{mi}(\mathbf{r})$ equals zero must be deleted from this set.
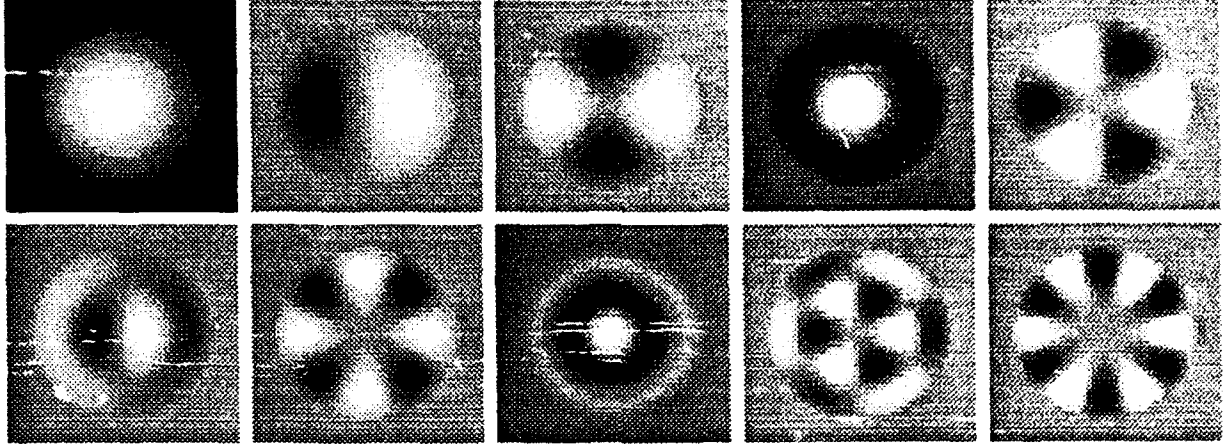
3

Figure 1: The shapes of the first ten principal components. ($w^1_{mi}(\mathbf{r})$ with $a = 1$ and $\phi_{mi} = 0$)

$$w^1_{mi}(\mathbf{r}) = \begin{cases} J_m(\frac{r}{\sqrt{\lambda_{mi}}})cos(m\theta + \phi_{mi}) & \text{for } r \leq a \\ 0 & \text{for } r > a \end{cases}$$

$$w^2_{mi}(\mathbf{r}) = \begin{cases} J_m(\frac{r}{\sqrt{\lambda_{mi}}})sin(m\theta + \phi_{mi}) & \text{for } r \leq a \\ 0 & \text{for } r > a \end{cases}$$

(5)

where ϕ_{mi} is a set of undetermined phases. These two eigen-functions have the same eigen-value λ_{mi}, i.e., they are degenerate.

If we order the solutions by the magnitudes of the correspondent eigenvalues λ_{mi}, the first ten solutions, $w^1_{mi}(\mathbf{r})$ with $\phi_{mi} = 0$ and $a = 1$, are drawn in figure 1.

3 Retrieving the Phase

The solutions above $w^1_{mi}(\mathbf{r})$ and $w^2_{mi}(\mathbf{r})$ not only have undetermined phases, but also are degenerate. This contradicts the results of the simulations (Hancock et al., 1992) in which the phases seem to always take the value zero, and the w^1_{mi} solution has a different eigenvalue from the w^2_{mi} solution. These results can be retrieved if we assume that the covariance matrix has a non-rotationally symmetric perturbation term. This assumption is not arbitrary since an inspection of Fields results(Field, 1987, figure 7) reveals that this is indeed the case. Hereafter we assume this perturbation term has, in $\mathbf{k}$ space, the form

$$C'(\mathbf{k}) = U(k)T(\theta_{\mathbf{k}}). \tag{6}$$

In order to calculate this perturbation, the representation of this perturbation in the two degenerate eigen-functions $w^1_{mi}(\mathbf{r})$ and $w^2_{mi}(\mathbf{r})$ has to be calculated. It is easier to perform

this in **k** space in which the eigen-functions $w_{mi}^1(\mathbf{r})$ and $w_{mi}^2(\mathbf{r})$ are replaced by their Fourier transforms,

$$w_{mi}^1(\mathbf{k}) = f_{mi}(k)cos(m\theta_{\mathbf{k}} + \phi_{mi})$$

$$w_{mi}^2(\mathbf{k}) = f_{mi}(k)sin(m\theta_{\mathbf{k}} + \phi_{mi})$$

(7)

in which

$$f_{mi}(k) = \pi j^m \int_0^a J_m(\frac{r}{\sqrt{\lambda_{mi}}}) J_m(kr) r dr$$

(8)

where $j^2 = -1$. If we denote

$$T(\theta_{\mathbf{k}}) = \sum_l t_l cos(l(\theta_{\mathbf{k}} - \alpha_l))$$

(9)

which is the Fourier expansion of $T(\theta_{\mathbf{k}})$. The representation of the perturbation matrix with respect to the two degenerate eigen-functions has the form

$$
\begin{aligned}
(C'_{(\mu,m,i|\gamma,m,i)})_{(\mu=1,2|\gamma=1,2)} &= (\int w_{mi}^\mu(\mathbf{k})^* C'(\mathbf{k}) w_{mi}^\gamma(\mathbf{k}) d^2\mathbf{k})_{(\mu=1,2|\gamma=1,2)} \\
&= g_{mi} \begin{pmatrix} cos(\delta) & sin(\delta) \\ sin(\delta) & -cos(\delta) \end{pmatrix}
\end{aligned}
$$

(10)

in which $\delta = 2\phi_{mi} + 2m\alpha_{2m}$ and $g_{mi} = \frac{\pi}{2} t_{2m} \int U(k) |f_{mi}(k)|^2 k dk$. Since the two eigen-functions are degenerate, any linear combination of these two eigen-functions is an eigen-function of C. Therefore, all we have to do is to find a linear combination of them which diagonalizes the perturbation matrix, i.e., to find the eigenvalues and eigen-vectors of the matrix in equation 10, which are

$$\begin{pmatrix} cos(\delta/2) \\ sin(\delta/2) \end{pmatrix}$$

(11)

$$\begin{pmatrix} -sin(\delta/2) \\ cos(\delta/2) \end{pmatrix}$$

with eigenvalues g_{mi} and $-g_{mi}$ respectively. Furthermore, if $U(k) = \epsilon/k^2$ then the complete expression for the correction to the eigenvalue takes the form $g_{mi} = \epsilon \lambda_{mi} t_{2m}/2$.

Thus the eigen-functions and eigenvalues after the perturbation can be readily written out as

$$w_{mi}^+(\mathbf{k}) = J_m(\frac{r}{\sqrt{\lambda_{mi}}}) cos(m(\theta - \alpha_{2m}))$$

(12)

$$w_{mi}^-(\mathbf{k}) = J_m(\frac{r}{\sqrt{\lambda_{mi}}}) sin(m(\theta - \alpha_{2m}))$$

5

with eigenvalues $\lambda_{mi}^+ = \lambda_{mi} + g_{mi}$, and $\lambda_{mi}^- = \lambda_{mi} - g_{mi}$, respectively. So the degeneracy is broken. This is in agreement with Hancock's simulations. These solutions have an important feature, i.e., their phases are determined by the properties of the real world covariance matrix. If the covariance matrix has a definite symmetry with an inclination angle α, then the solutions would also have the same symmetry angle. Because in this case $\alpha_{2m} = \alpha$ for all m. The spectrums of the covariance matrix, shown in figure 7 of Field's paper, indeed indicates a symmetry axis along $\alpha = 0$. Thus equation 12 predicts the zero phase result found in Hancock's simulation. When Hancock used images which were tilted by 45 degrees before being scanned, the preferred axis of the receptive fields was found to be 45 degrees. Again this is predicted by equation 12, because the symmetry axis of the covariance matrix spectrum also gets rotated by 45 degrees due to the rotated images, i.e., $\alpha = 45°$, and thus the solutions also get rotated by 45 degrees.

4   Discussion

We have calculated the forms of the principal components of natural images based on the result about the covariance matrix (Field, 1987), and have shown that a non-rotationally symmetric perturbation can break the degeneracy and give a definite phase which only depends on the properties of the real world covariance matrix. These results for a large part agree with the numerical simulation(Hancock et al., 1992).

A discrepancy can be noticed between our results and Hancock's simulation results for eigen-functions corresponding to smaller eigenvalues. This comes from the mixing of solutions, due to a perturbation added to the covariance matrix(Merzbacher, 1970). This mixing would be more noticeable for lower eigenvalue solutions since for smaller λ_{mi} and $\lambda_{m'i}$, i.e., for large m and m', $\lambda_{mi} \approx \lambda_{m'i}$, the solutions are nearly degenerate. Thus any small perturbation will yield highly mixed linear combinations of different eigen-functions, the same linear combination phenomena we have discussed in section 3. The perturbation can come from the difference between the Gaussian mask used in the simulation, and the hard boundary we chose in the analytic analysis. It can also come from all sorts of correction to the spectrum of the covariance matrix including the non-rotationally symmetric correction. Another correction can be a high frequency cutoff due to the finite sampling frequency of the images.

The neurobiological relevance of the type of technique used in this paper is that we can deduce for different learning rules what kinds of receptive fields they should produce. Given these receptive fields, we can compare them to the real biological receptive fields. This comparison can be used to assess whether the biological hardware really implements or approximates a theoretically proposed learning rule.

The most obvious conclusion which stands out when we observe the results in figure 1, is that these receptive fields have little resemblance to receptive fields reported in the biological literature (Hubel and Wiesel, 1959, etc.). Does this imply that biological neurons are not principal component analizers? When addressing this question we have to keep in mind that the natural images projected on the retina, undergo preprocessing in the retina and LGN,

6

before they reach the visual cortex. Similar preprocessing should therefore be applied to natural images in simulations and analytic studies, before a sensible answer can be given. What can also be observed is that the receptive fields of the neurons in visual cortex seem to form a non-orthogonal set such as Gabor function family (Daugman, 1985, etc.) as opposed to the orthogonal and complete function set we got here. This may result from neurons being noisy entities, a property of neurons not being taken into account when formulating the minimal mean square reconstruction error criterion, which results in the principal component eigen-equation. All these issues are currently under study.

5 Acknowledgments

References

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2:1160–1169.

Field, D. J. (1987). Ralations between the statistics of natural images and the response properities od cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379–2394.

Fukunaga, K. (1990). *Introduction to Statitical Pattern Recognition (2rd Ed.)*. Academic Press, Inc.

Hancock, P. J., Baddeley, R. J., and Smith, L. S. (1992). The principal components of natural images. *Network*, 3:61–70.

Hebb, D. O. (1949). *The Organization of Behavior*. Wiley.

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurons in the cat strate cortex. *J. Physiol. (London)*, 148:509–591.

Jackson, J. D. (1975). *Classical Electrodynamics*. Wiley.

Kammen, D. and Yuille, A. (1988). Spontaneous symmetry-breaking energy function and the emergence of selective cortical cells. *Bio. Cybern.*, 59:23–31.

Linsker, R. (1986). From basic network principles to neural architecture. *Proc. Natl. Acad. Sci. USA*, 83:7508–12, 8390–4, 8779–83.

Linsker, R. (1988). Self-organization in a perceptual network. *Computer*, pages 105-17.

MacKay, D. J. and Miller, K. D. (1990a). Analysis of linsker's simulation of hebbian rules. *Neural Computation*, 1:173–187.

MacKay, D. J. and Miller, K. D. (1990b). Analysis of linsker's simulations of hebbian rules to linear networks. *Network*, 1:257–297.

Merzbacher, E. (1970). *Quantum Mechnics*. Wiley.

Oja, E. (1982). A simplified neuron model as a principal component ananlyzer. *J. Math. Bio.*, 15:267–273.

Sanger, T. D. (1989). Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2:459–473.

Tang, D. S. (1989). Information-theoretic solutions to early visual information processing: Analytic results. *Phys. Rev. A*, 40(11):6626–35.

Yuille, A., Kammen, D., and Cohen, D. (1989). Quadrature and the development of orientation selective cells by hebb rules. *Bio. Cybern.*, 61:183–194.